

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number  
**WO 01/01218 A2**

- (51) International Patent Classification<sup>7</sup>: G06F Andreas, K. [DE/US]; 91 Center Road, Woodbridge, CT 06524 (US). XU, Chuanbo [CN/US]; 524 Openint Hill Road, Madison, CT 06443 (US).
- (21) International Application Number: PCT/US00/17540
- (22) International Filing Date: 26 June 2000 (26.06.2000) (74) Agents: MOROZ, Eugene et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/141,521 25 June 1999 (25.06.1999) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/141,521 (CIP)  
Filed on 25 June 1999 (25.06.1999)
- (71) Applicant (for all designated States except US): GENAIS-SANCE PHARMACEUTICALS, INC. [US/US]; Five Science Park, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): DENTON, Richard, Rex [US/US]; 129 Hunters Trail, Madison, CT 06443 (US). JUDSON, Richard, S. [US/US]; 42 Baker Hill Drive, Guilford, CT 06437 (US). RUAÑO, Gualberto [US/US]; 88 Lawrence Street, New Haven, CT 06511 (US). STEPHENS, Joel, Claiborne [US/US]; 46 Crabapple Lane, Guilford, CT 06437 (US). WINDEMUTH,
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— Without international search report and to be republished upon receipt of that report.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/01218 A2

(54) Title: METHODS FOR OBTAINING AND USING HAPLOTYPE DATA

(57) Abstract: Methods, computer program(s) and database(s) to analyze and make use of gene haplotype information. These include methods, program, and database to find and measure the frequency of haplotypes in the general population; methods, program, and database to find correlation's between an individual's haplotypes or genotypes and a clinical outcome; methods, program, and database to predict an individual's haplotypes from the individual's genotype for a gene; and methods, program, and database to predict an individual's clinical response to a treatment based on the individual's genotype or haplotype.

- 36 -

reference population is that described in U.S. Application Serial No. 60/198,340 (inventors Stephens et al.), filed April 18, 2000, which is specifically incorporated by reference herein. Another, less preferred embodiment for haplotyping the reference population, uses the CLASPER System™ technology (Ref. U.S. Patent Number 5,866,404), which is a technique for direct haplotyping. Other examples of the techniques for direct haplotyping include single molecule dilution ("SMD") PCR (Ref. 9) and allele-specific PCR (Ref. 10). However, for the purpose of this invention, any technique for producing the haplotype information may be used.

The information that is stored in a database, such as a database associated with the DecoGen application exemplified herein includes (1) the positions of one or more, preferably two or more, most preferably all, of the sites in the gene locus (or other loci) that are variable (i.e. polymorphic) across members of the reference population and (2) the nucleotides found for each individuals' 2 haplotypes at each of the polymorphic sites. Preferably, it also includes individual identifiers and ethnicity or other phenotypic characteristics of each individual.

In the preferred embodiment of the invention, the haplotypes and their frequencies are stored and displayed, preferably in the manner shown, e.g., in FIGURES 4A and 4B. Haplotypes and other information about each of the members of the population being analyzed can be shown, for example, in the manner shown in FIGURE 8. The information shown in FIGURE 8 includes a unique identifier (PID), ethnicity, age, gender, the 2 haplotypes seen for the individual, and values of all clinical measurements available for the individual. Quantitative values of clinical measures would ordinarily be seen by scrolling to the right. However, for the subjects seen in this view, there is no clinical data. This is because this is the reference population of healthy individuals.

The haplotype data may also be presented in the context of the entire DNA sequence. Examples of the sequences of the isogenes, with the polymorphisms highlighted, are shown in FIGURE 5.

Because an individual has 2 copies of the gene (2 isogenes), and because these 2 copies are often different, some of the polymorphic sites will show 2 different nucleotides in a genotype, one from each of the isogenes. A genotype from an individual with haplotypes TAC and CAG would be

- 23 -

° chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

**Genotyping** – A process for determining a genotype of an individual.

5 **Haplotype** – A member of a polymorphic set, e.g., a sequence of nucleotides found at one or more of the polymorphic sites in a locus in a single chromosome of an individual. (See, e.g., HAP 1 in FIGURE 4A full haplotype is a member of a full polymorphic set). A sub-haplotype is a member of a polymorphic subset.

10 **Haplotype data** – Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or  
15 more haplotypes and a trait.

**Haplotype pair** – The two haplotypes found for a locus in a single individual.

**Haplotyping** – A process for determining one or more  
20 haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

**Isoform** – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

25 **Isogene** – One of the two copies (or isoforms) of a gene possessed by an individual or one of all the copies (or isoforms) of the gene found in a population. An isogene contains all of the polymorphisms present in the particular copy (or isoforms) of the gene.

30 **Isolated** – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term  
35 "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that